Industrial Organization: Problem Set 2, The Redemption Tour

Tanya Rajan and George Vojta

February 27, 2021

Question 1.1: Ciliberto and Tamer (2008) Bounds

To get the bounds for $\mathbb{P}(y_{1t} = 1, y_{2t} = 0|x_t)$ we should first write out scenarios in which firm 1 enters but first two doesn't:

1. It is never profitable for Firm 2 to enter, and Firm 1 is in the range where it is profitable for them to enter when Firm 2 doesn't. So Firm 1 always enters.
2. Firm 1 always wants to enter, Firm 2 would like to enter but only if firm 1 does not and thus does not enter
3. Firm 1 would like to enter but only if firm 2 does not, Firm 2 would like to enter but only if firm 1 does not, there is some sort of random coin flip behavior.

We can define these portions similarly as:

1. $R_1(\delta, \alpha, x_it) = \{(\varepsilon_{1t}, \varepsilon_{2t})|\epsilon_{1t} \geq -\alpha x_t; \epsilon_{2t} \leq -\alpha x_t\}$
2. $R_2(\delta, \alpha, x_it) = \{(\varepsilon_{1t}, \varepsilon_{2t})|\epsilon_{1t} \geq -\alpha x_t + \delta; -\alpha x_t + \delta \ \geq \epsilon_{2t} \geq -\alpha x_t\}$
3. $R_3(\delta, \alpha, x_it) = \{(\varepsilon_{1t}, \varepsilon_{2t})| -\alpha x_t + \delta \geq \epsilon_{1t} \geq -\alpha x_t; -\alpha x_t + \delta \ \geq \epsilon_{2t} \geq -\alpha x_t\} \implies (\epsilon_{1t}, \epsilon_{2t}) \in [-\alpha x_t, -\alpha x_t + \delta]$

First let us define two relevant regions. The lower bound is the probability in the regions where Firm 1 will always enter (i.e. Regions R_1 and R_2). We are assuming that when both firms have a chance of entering, Firm 1 never enters the market (i.e. some tiebreak rule that always rules against Firm 1).

$$\mathcal{R}^L = R_1(\delta, \alpha, x, t) \cup R_2(\delta, \alpha, x, t)$$

The upper bound region of interest is where both firms might enter. So the region of interest is:

$$\mathcal{R}^U = R_3(\delta, \alpha, x, t)$$

The probability that Firm 1 enters can be exactly defined as:

$$\int \underbrace{\mathbb{1}\{(\varepsilon_1, \varepsilon_2) \in \mathcal{R}^L\}}_{\text{Firm 1 always enters}} + \int \underbrace{\mathbb{P}(y_{i1} = 1, y_{i0} = 0)}_{\text{Prob. only firm 1 enters}} \underbrace{\mathbb{1}\{(\varepsilon_1, \varepsilon_2) \in \mathcal{R}^U\}}_{\text{Both firms might enter}} dF_\varepsilon$$

We know that the probability $\mathbb{P}(y_{i1} = 1, y_{i0} = 0)$ is bounded between 0 and 1, so by setting its values to the extremes, we get our upper and lower bounds:

$$\int \mathbb{1}\{(\varepsilon_1, \varepsilon_2) \in \mathcal{R}^L\} dF_\varepsilon \leq \mathbb{P}(y_{1t} = 1, y_{2t} = 0|x_t) \leq \int \mathbb{1}\{(\varepsilon_1, \varepsilon_2) \in \mathcal{R}^L\} dF_\varepsilon + \int \mathbb{1}\{(\varepsilon_1, \varepsilon_2) \in \mathcal{R}^U\} dF_\varepsilon$$

1.2 Demand Estimation with Endogenous Entry

Question 1

We cannot get consistent estimates of β and γ from the OLS regression for two reasons. First, prices may be correlated with ξ_{it}, as we discussed in the BLP case. Second, we only observe market share data when a given firm decides to enter the market, so market share in the data is actually $S_{it}y_{it}$. Since ξ, ε are jointly distributed, there is the possibility that factors that affect firm entry also affect subsequent market share.

Question 2

A price instrument may be able to address the first issue above, as it does in the BLP setting. However, it will not necessarily address the selection bias issue due to the endogenous entry component.

Question 3

Again, let us define two regions of interest. First, define the region $A^U_{(1,0)}$, where $(y_{1t}, y_{2t}) = (1,0)$ is the unique equilibrium. Next, define the region A^M where $(1,0)$ is one of multiple possible equilibria. Then the probability of interest can be exactly defined as:

$$\begin{aligned}\mathbb{P}\left(\xi_{1t} \leq e | y_{1t} = 1, y_{2t} = 0, x, z\right) = \mathbb{P}\left(\xi_{1t} \leq e | \left(\varepsilon_1, \varepsilon_2\right) \in A^U_{(1,0)}, x, z\right) + \\ \mathbb{P}\left(d_{(1,0)} = 1 \mid \xi_1 \leq e, (\varepsilon_1, \varepsilon_2) \in A^M x, z\right) \mathbb{P}\left(\xi_{1t} \leq e \mid (\varepsilon_1, \varepsilon_2) \in A^M, x, z\right)\end{aligned}$$

where $d_{(1,0)}$ represents the equilibrium (1,0) being chosen. We know that the probability of $(1,0)$ being chosen is bounded between 0 and 1. So we can get our upper and lower bounds by evaluating the equation above when the probability of choosing (1,0) as the equilibrium is at either extreme:

$$\begin{aligned}\mathbb{P}(\xi_{1t} \leq e | (\varepsilon_1, \varepsilon_2) \in A^U_{(1,0)}, x, z) &\leq \mathbb{P}(\xi_{1t} \leq e, y_{1t} = 1, y_{2t} = 0, x, z) \\ &\leq \mathbb{P}(\xi_{1t} \leq e | (\varepsilon_1, \varepsilon_2) \in A^U_{(1,0)}, x, z) + \mathbb{P}\left(\xi_{1t} \leq e \mid (\varepsilon_1, \varepsilon_2) \in A^M, x, z\right)\end{aligned}$$

1.3 Incomplete Information

Question 1

Since firms only observe their own idiosyncratic shock, we can now rewrite firm i's optimal entry strategy as:

$$y_{it} = \mathbb{1}\{-\delta \mathbb{E}[y_{-i,t}|x_t] + \alpha x_t + \varepsilon_{it} \geq 0\}$$

Define $P_{i,t}(x_t) = \mathbb{E}[y_{i,t}|x_t]$. Taking expectations of the equation above, we get:

$$P_{it}(x_t) = \mathbb{P}(\delta P_{-i,t} - x_t\alpha \leq \varepsilon_{it}|x_t) = 1 - F_\varepsilon(\delta P_{-i,t} - x_t\alpha)$$

We know that $\varepsilon \sim Logistic(0,1)$, so we can the behavior of firm i from the other firm's point of view as:

$$P_{it} = \frac{\exp\left(x_t\alpha - \delta P_{-i,t}\right)}{1 + \exp\left(x_t\alpha - \delta P_{-i,t}\right)}$$

for $i \in \{1,2\}$. So we have two interdependent equations (one per firm) that we can solve to find BNEs.

Question 2

We use Newton's method with tolerance set at 10e-12 to solve the fixed point equations from above (see code and writeup below for details). We solve the equations for a grid of starting values on $[0,1] \times [0,1]$ and find the following equilibria:

For $x = 1, (\alpha, \delta) = (1, 1)$: (0.5989, 0.5989)
For $x = 2, (\alpha, \delta) = (1, 1)$: (0.7732, 0.7732)
For $x = 1, (\alpha, \delta) = (3, 6)$: (0.0707, 0.9293), (0.5, 0.5), (0.9293, 0.0707)
For $x = 2, (\alpha, \delta) = (3, 6)$: (0.7118, 0.8493), (0.7846, 0.7846), (0.8493, 0.7118)

Note that we use the multivariate form of Newton's method. Define $f^i = \frac{\exp(x_t\alpha - \delta P_{-i,t})}{1+\exp(x_t\alpha - \delta P_{-i,t})} - P_{it}$ for $i \in \{1, 2\}$. Then write $f = (f^1, f^2) : \mathbb{R}^2 \to \mathbb{R}^2$ and $P = (P_1, P_2)$. Newton's method gives us:

$$P' = P - J_P^{-1} f(P)$$

where J_P is the Jacobian of f evaluated at the point P. Taking derivatives of the functions f^i, we can solve for elements of this Jacobian:

$$\begin{aligned}
J_{11} &= \frac{\partial f^1}{\partial P_1} = -1 \\
J_{12} &= \frac{\partial f^1}{\partial P_2} = \frac{-\delta \exp(x_t\alpha - \delta P_2)}{(1 + \exp(x_t\alpha - \delta P_2))^2} \\
J_{21} &= \frac{\partial f^2}{\partial P_1} = \frac{-\delta \exp(x_t\alpha - \delta P_1)}{(1 + \exp(x_t\alpha - \delta P_1))^2} \\
J_{22} &= \frac{\partial f^2}{\partial P_2} = -1
\end{aligned}$$

Question 3

The likelihood function for $\omega = (\alpha, \delta)$ is

$$\begin{aligned}
\mathcal{L} &= \prod_{t=1}^{T}\prod_{i=1}^{2} \mathbb{P}\left(y_{it} = 1 \mid x_t, P_{-it}, \omega\right)^{y_{it}} \left\{1 - \mathbb{P}\left(y_{ir} = 1 \mid x_t, P_{-it}, \omega\right)\right\}^{1-y_i} \\
\ln \mathcal{L} &= \sum_{t=1}^{T}\sum_{i=1}^{2} y_{it} \ln \mathbb{P}\left(y_{it} = 1 \mid x_t, P_{-i,t}, \omega\right) + (1 - y_{it}) \ln\left(1 - \mathbb{P}\left(y_{i+} = 1 \mid x_t, P_{-i,t}, \omega\right)\right)
\end{aligned}$$

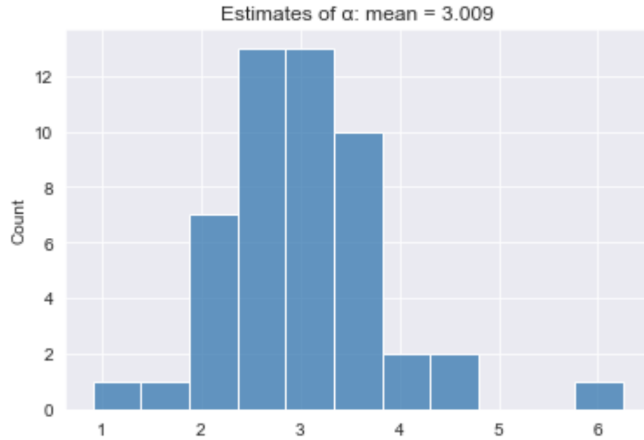
Note that we can rewrite the probability terms as follows:

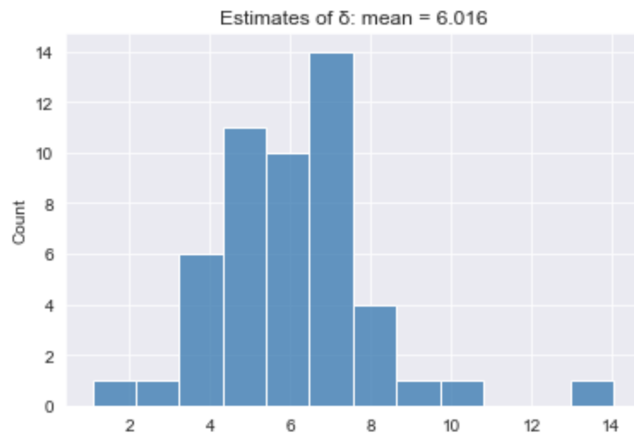
$$\begin{aligned}
\mathbb{P}\left(y_{it} = 1 \mid x_t, P_{i,t}, \omega\right) &= \mathbb{P}\left(-\delta P_{-i,t} + \alpha x_t + \varepsilon_{it} \geqslant 0\right) \\
&= \mathbb{P}\left(\varepsilon_{it} \geqslant \delta P_{-i,t} - \alpha x_t\right) \\
&= 1 - F_{\varepsilon}\left(\delta P_{-i,t} - \alpha x_t\right) \\
&= \frac{\exp(g_{-i,t})}{1 + \exp(g_{-i,t})}
\end{aligned}$$

where $g_{-i,t} := \alpha x_t - \delta P_{-i,t}$. For symmetric strategies, $P_{-i,t}$ is the same for both firms, so we can write $g_{-i,t} = g_t$. Then the log likelihood is:

$$\ln \mathcal{L} = \sum_{t=1}^{T}\sum_{i=1}^{2} y_{it} \ln\left(\frac{\exp(g_t)}{1 + \exp(g_t)}\right) + (1 - y_{it}) \ln\left(\frac{1}{1 + \exp(g_t)}\right)$$

The estimates from 50 Monte Carlo samples of $T = 1000$ observations are plotted below. They are roughly centered around $\alpha = 3, \delta = 6$, as we would expect. The standard deviations are 0.8376 and 2.0132, respectively.





Question 4

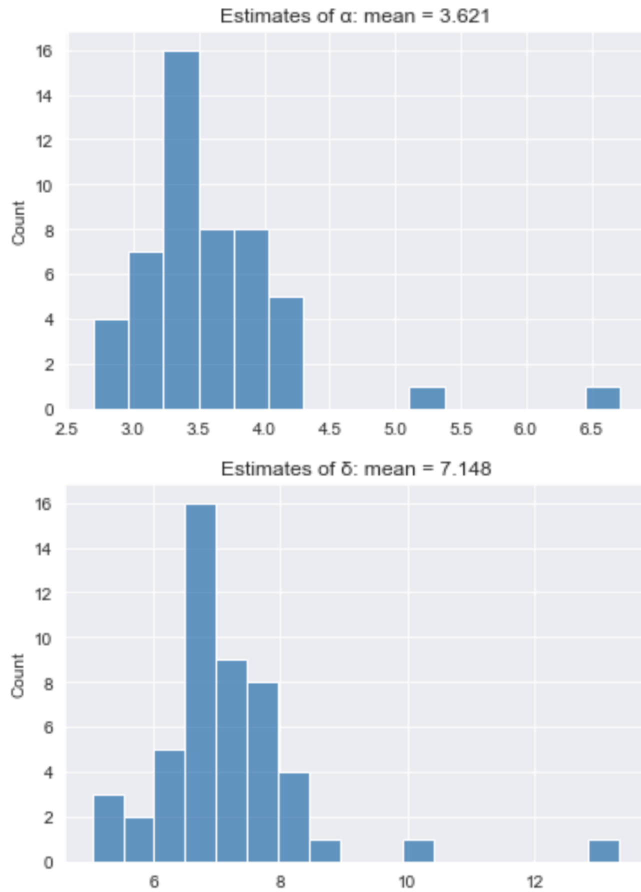
Now we have multiple equilibria, $k = 1, \ldots, K$, indexed in descending order by firm 1's choice probability $P^k_{1,t}$, that can be chosen with probability $\lambda^k_{it}(x_t) \sim \exp\left(\frac{k}{2}\right)$. We can rewrite the likelihood function as follows. Define $\bar{y}_i = (y_{i1}, \ldots, y_{iT})$ and $\bar{x} = (x_1, \ldots, x_T)$.

$$\begin{aligned}
\mathbb{P}(\bar{y}_1, \bar{y}_2 | \bar{x}, \omega) &= \prod_{t=1}^{T} \mathbb{P}(y_{1t}, y_{2t} | x_t, \omega) \\
&= \prod_{t=1}^{T} \prod_{i=1}^{2} \mathbb{P}(y_{it} | x_t, \omega) \\
&= \prod_{t=1}^{T} \prod_{i=1}^{2} \sum_{k=1}^{K} \mathbb{P}(y_{it} | x_t, \omega, P^k_{it}) \mathbb{P}(\text{select } k | x_t)
\end{aligned}$$

where $\mathbb{P}(\text{select } k|x_t)$ is the probability that equilibrium k is chosen, which is equal to the exponential PDF: $\lambda \exp(-\lambda x)$ defined on $[0, \infty)$, with parameter $\lambda = \frac{k}{2}$. The last line follows from the law of total probability and the fact that choice of equilibriium affects entry decisions only through P^k_{it}. We already derived the log likelihood of the term $\mathbb{P}(y_{it}|x_t, \omega, P^k_{it})$ above. So we can rewrite this in log likelihood terms as:

$$\ln \mathcal{L}^* = \sum_{t=1}^{T} \sum_{i=1}^{2} \ln \left(\sum_{k=1}^{K} \frac{k}{2} \exp\left(-\frac{k}{2} x_t\right) \left(\frac{\exp\left(g^k_{-i,t}\right)}{1 + \exp\left(g^k_{-i,t}\right)} \right)^{y_{it}} \left(\frac{1}{1 + \exp\left(g^k_{-i,t}\right)} \right)^{(1-y_{it})} \right)$$

where $g^k_{-i,t} := \alpha x_t - \delta P^k_{-i,t}$. We propose $\hat{\omega}_{MLE} = \arg\max \ln \mathcal{L}^*$ as a consistent estimator. The Monte Carlo plots below show the distribution of estimates. Standard deviations are 0.619 for α and 1.2341 for δ The estimates are not as clearly centered around the true mean as in the previous part. However the true values 3 and 6 are within one standard deviation of the estimated means for $\hat{\alpha}$ and $\hat{\delta}$.



Question 2.3

Part 1

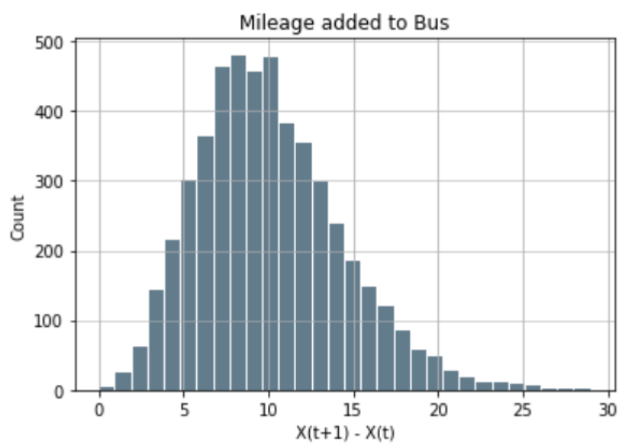
You can get engine replacement decisions whenever the mileage is less than it was in the last period. It is possible that for some reason at very low mileages that our friend Harold Zurcher actually replaces the engine and then proceeds to drive further than all the mileage that was previously on the bus in one period, but in all likelihood this won't happen so we'll just assume that problem away.

Part 2

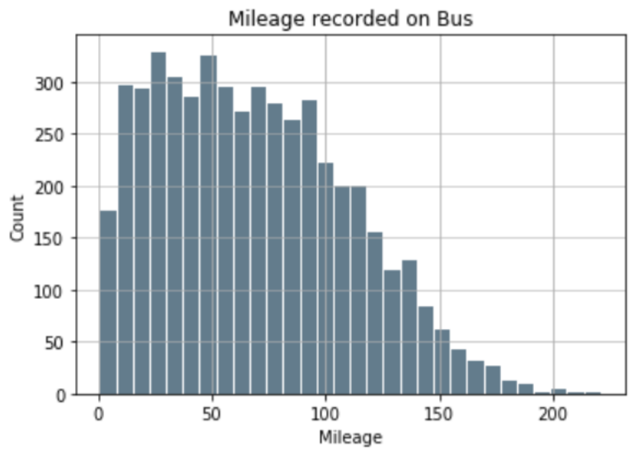
The conditional independence assumption states that given (x_t, a_t), we know that $x_{t+1} \perp (\varepsilon_t, \varepsilon_{t+1})$ which implies $p(x_{t+1}, \varepsilon_{t+1}|x_t, \varepsilon_t, a_t) = \rho(x_{t+1}|x_t, a_t)g(\varepsilon_{t+1})$. So ε_t only affects x_{t+1} through the choice of action a_t. This means that these utility shocks are orthogonal to how the state may change (conditioning on choice a_t). However, we can think of something like a summer road trip season where people gain additional utility from choosing to go on a trip and also drive for longer distances in the bus. So the state will change differently since buses are racking up higher mileages, while the utility function is also influenced by the same shock.

Part 3

Below we plot the histogram of added miles to bus per period:



The added increment was created by subtracting off the last period mileage of the bus, unless the bus had undergone an engine change, at which point ΔMileage = Mileage. This can be thought of as our distribution for the $g(.)$ term. Note that the range of $g(.)$ in the data is 0-29. Now lets plot the overall bus mileage histogram:



We discretize the state space into $K = 10$ bins of equal width (in terms of mileage). We then construct the transition matrix by taking each bin (b_t) and computing the frequency of observing observations moving from bin b_t to each bin $b_{t+1} \in \{1, \ldots, K\}$. We get a transition matrix with rows that sum to 1 as desired (displayed below) See the code for additional details.

```
array([[0.569, 0.43 , 0.001, 0.   , 0.   , 0.   , 0.   , 0.   , 0.   ,
        0.   ],
       [0.   , 0.569, 0.43 , 0.001, 0.   , 0.   , 0.   , 0.   , 0.   ,
        0.   ],
       [0.   , 0.   , 0.569, 0.43 , 0.001, 0.   , 0.   , 0.   , 0.   ,
        0.   ],
       [0.   , 0.   , 0.   , 0.569, 0.43 , 0.001, 0.   , 0.   , 0.   ,
        0.   ],
       [0.   , 0.   , 0.   , 0.   , 0.569, 0.43 , 0.001, 0.   , 0.   ,
        0.   ],
       [0.   , 0.   , 0.   , 0.   , 0.   , 0.569, 0.43 , 0.001, 0.   ,
        0.   ],
       [0.   , 0.   , 0.   , 0.   , 0.   , 0.   , 0.569, 0.43 , 0.001,
        0.   ],
       [0.   , 0.   , 0.   , 0.   , 0.   , 0.   , 0.   , 0.569, 0.43 ,
        0.001],
       [0.   , 0.   , 0.   , 0.   , 0.   , 0.   , 0.   , 0.   , 0.569,
        0.431],
       [0.   , 0.   , 0.   , 0.   , 0.   , 0.   , 0.   , 0.   , 0.   ,
        1.   ]])
```

Part 4

Starting from:

$$EV(x, d) = \int V_\theta * y, \epsilon) p(d\epsilon) p(dy|x, d)$$

Let's start by decomposing V_θ into the the choice of whether $d = a, a \in \{0, 1\}$:

$$\begin{aligned} V_\theta(x_t, \epsilon_{a,t}) &= \max\{\text{replace}, \text{don't replace}\} \\ &= \max_{a \in \{0,1\}} \{u(x_t, d_t = a) + \varepsilon_{a,t} + \beta \int_y \int_{\epsilon_{t+1}} v(y, \epsilon_{t+1}) p(d\epsilon) p(dy|x, d = a)\} \end{aligned}$$

Now we can define $w(y) = \int_{\epsilon_{t+1}} v(y, \epsilon_{t+1}) p(d\epsilon)$ and write:

$$\begin{aligned} V_\theta(x_t, \epsilon_{a,t}) &= \max\{\text{replace}, \text{don't replace}\} \\ &= \max_{a \in \{0,1\}} \{u(x_t, d_t = a) + \varepsilon_{a,t} + \beta \int_y w(y) p(dy|x, d = a)\} \end{aligned}$$

Now using our definition of V_θ we can plug this BACK into our $w(y)$ expression to get a recurrsive expression:

$$w(x_t) = \int_{\epsilon_{t+1}} \left(\max_a \{u(x_t, d_t = a) + \varepsilon_{a,t} + \beta \int_y w(y)p(dy|x, d = a)\} \right)$$

And define $\delta_a := u(x_t, d = a) + \beta \int_y w(y)p(dy|x, d = a)$ to separate out the epsilon components

$$\begin{aligned}
w(x) &= \int_{\epsilon_{t+1}} \left(\max_{a \in \{0,1\}} \{\delta_a + \epsilon_j\} \right) p(d\epsilon) \\
&= log \left(\sum_a exp(\delta_a) \right) + k \\
\Rightarrow w(x, \theta) &= log \left(\sum_a exp(u(x_t, d_t = a) + \beta \underbrace{\int_y w(y)p(dy|x, d_t = a))}_{:=EV(y, d_t = a)} \right) + k
\end{aligned}$$

Plugging this all in we get:

$$\begin{aligned}
EV(x_t, d_t) &= \int_{\epsilon_{t+1}} w(y)p(dy|x_t, d_t) \\
&= \int_{\epsilon_{t+1}} \underbrace{log \left(\sum_a exp(u(x_t, d_t = a) + \beta EV(y, d_t = a) \right)}_{\Gamma(EV)} p(dy|x_t, d_t)
\end{aligned}$$

Part 5

We can define the conditional probabilities as:

$$\begin{aligned}
\mathbb{P}(\text{replace}|x_t, \theta) &= \mathbb{P}(V_\theta(x_t, 1) > V_\theta(x_t, 0)) \\
&= \mathbb{P}(V_\theta(x_t, 1) - V_\theta(x_t, 0) \geq \varepsilon_{1,t} - \varepsilon_{0,t}) \\
&= \frac{1}{1 + \exp(V_\theta(x_t, 0) - V_\theta(x_t, 1))} \\
&= \frac{V_\theta(x_t, 1)}{\sum_{d \in \{0,1\}} V_\theta(x_t, 1)} \\
&= \frac{exp\left(u(x_t, d_t = 1, \theta) + \beta \int_y log[\sum_a exp(u(x_t, d_t = a) + \beta EV(y, d_t = 1)]\right)}{\sum_a \left(exp\left(u(x_t, d_t = a, \theta) + \beta \int_y log[\sum_a exp(u(x_t, d_t = a) + \beta EV(y, d_t = a)]\right)\right)}
\end{aligned}$$

The third line comes from the T1EV property that we derived on the previous problem set: if $u_j \sim T1EV(\delta_j), u_k \sim T1EV(\delta_k)$, then $u_j - u_k \sim Logistic(\delta_j - \delta_k))$. The probability of replacement can be defined similarly or as: $\mathbb{P}(\text{don't replace}|x_t, \theta) = 1 - \mathbb{P}(\text{replace}|x_t, \theta)$

Question 2.3.1

Part 1

Because replacing the engine automatically sets the $EV(x, d = 1) = \text{Constant} = K$ due to the fact it does not depend on this periods x_t allows us to considerably reduce the state space. We can set $EV(x_t, d = 1) = EV(0, d = 0)$, so we can now write all the value functions in terms of the state variable x_t alone.

Part 2

Because we are integrating over the transitions to the next mileage (state variable), we can think of our integral as just multiplying by these:

$$EV(x,d) = K_d \cdot log\left(\sum [exp(u(x_t,d_t) + \beta EV(y,d_t))]\right)$$

Were $U = U_0 + U_1$ are K dimensional column vectors, U_1 being the same repeated value over and over again and U_0 being a mapping from $x_t \Rightarrow \mathbb{R}$, EV is defined similarly. Thus the fixed point equation is between:

$$\begin{aligned} f^d &= EV(x,d) - K_d \cdot log\left(\sum [exp(u(x_t,d_t) + \beta EV(y,d_t))]\right) \\ &= (I - \Gamma)(EV(x,d)) \end{aligned}$$

Part 3

Please see attached code.

Part 4

The full likelihood is

$$\mathcal{L} = \prod_{t=1}^{T} \left(\prod_{a \in \{0,1\}} \mathbb{P}(d_t = a | x_t, \theta)^{\mathbb{1}\{d_t = a\}} | \right) \mathbb{P}(x_t | x_{t-1}, d_{t-1})$$

. Because the transition matrix term on the outside is independent of θ and becomes additively separable in the log likelihood form, we can focus instead on just the portion inside the parentheses to get consistent estimates. The likelihood can be mapped by the decision Zurcher makes on whether or not to replace his bus every period, which is a Bernoulli Distribution:

$$l(\theta) = \sum_t d_t log(1 - p_{x_t}(\theta)) + (1 - d_t) log(p_{x_t}(\theta))$$

Where p_{x_t} is the probability of *not* replacing the bus, which is:

$$p_k = \frac{1}{1 + exp(u(k,1|\theta) + \beta EV(0|\theta) - u(k,1|\theta) - \beta EV(k|\theta)}$$

Part 5

Sadly, while our likelihood is a nice looking function, it's derivative.... is atrocious. We'll do our best out here though. For ease let's start by defining:

$$\begin{aligned} V0 &:= u(k,0|\theta) + \beta EV(k|\theta) \\ V1 &:= u(k,1|\theta) + \beta EV(0|\theta) \end{aligned}$$

Using the chain rule on our derivative:

$$\frac{\partial l(\theta)}{\partial \theta} = \sum_t \left[\frac{d_t}{1 - p_k(\theta)} \left(\frac{-\partial p_k(\theta)}{\partial \theta} \right) + \frac{1 - d_t}{p_k(\theta)} \left(\frac{\partial p_k(\theta)}{\partial \theta} \right) \right]$$

Clearly, the key here is figuring out what $\frac{\partial p_k(\theta)}{\partial \theta}$ is:

$$\frac{\partial p_k(\theta)}{\partial \theta} = \frac{-\left[\frac{\partial u(k,1)}{\partial \theta} + \beta \frac{\partial EV(0)}{\partial \theta} - \frac{\partial u(k,0)}{\partial \theta} - \beta \frac{\partial EV}{\partial \theta} \right]}{(1 + exp(V1 - V0))^2}$$

It'll actually be easier to write this as:

$$p_k = \frac{exp(V0)}{exp(V0)+exp(V1)} = \frac{1}{1+exp(V1-V0)}$$
$$\Rightarrow (1-p_k) = \frac{exp(V1)}{exp(V0)+exp(V1)} = \frac{exp(V1-V0)}{1+exp(V1-V0)}$$
$$\Rightarrow p_k(1-p_k) = \frac{exp(V1-V0)}{(1+exp(V1-V0))^2}$$

Which allows us to simplify to:

$$\frac{\partial p_k(\theta)}{\partial \theta} = -p_k(1-p_k)\left[\frac{\partial u(k,1)}{\partial \theta} + \beta\frac{\partial EV(0)}{\partial \theta} - \frac{\partial u(k,0)}{\partial \theta} - \beta\frac{\partial EV}{\partial \theta}\right]$$

And this is considerably simpler than what we had previously. Now we have to go another step further and write out the following derivatives:

$$\frac{\partial u(k,1)}{\partial \theta} = \left[-x, -\left(\frac{x}{100}\right)^2, 0\right]$$
$$\frac{\partial u(k,0)}{\partial \theta} = [0,0,-1]$$
$$\frac{\partial EV}{\partial \theta} = (I-\Gamma')^{-1}\frac{\partial \Gamma}{\partial \theta}$$

Where the last line is from the slide 48 of the February 11th slides, and the first two derivatives come from trivial derivations. Finally we have to deal with the tricky $\frac{\partial \Gamma}{\partial \theta}$ term:

$$\frac{\partial \Gamma}{\partial \theta} = F_0 \frac{exp(V1)\frac{\partial u(k,1)}{\partial \theta} + exp(V0)\frac{\partial u(k,0)}{\partial \theta}}{exp(V1)+exp(V0)}$$

Great, now that we have typed this all out, we can code it up.

Part 6

In our main code (`pset2_q2.ipynb`), we put it all together and attempt to run the optimizer. However the results we get may suggest a bug in the code. The Rust Polyalgorithm for the first expression makes sense, with EV decreasing in X. However as the maximizer goes through the process this flips and its unclear. Running it through the likelihood we get:

$$\theta = \begin{bmatrix} 585.02 & -3232.13 & 0.15 \end{bmatrix}$$

However the output of the optimizer is really finnicky depending on the starting guess for our θ and doesn't make that much sense. At this point we believe problem is in the likelihood or it's gradient since the Rust Poly Algorithm gives legitimate results until it gets fed into there. It would be great if we could get feed back on where it went wrong but other wise this is as good as we can get. Honestly there's a chance the problem is not with the code but with the wrapper we attempted.

In an alternative attempt at the problem (see `pset2_q2_alt.ipynb`), we instead use `scipy`'s default BFGS method without providing a gradient function. In this case we are able to get estimates that seem slightly more reasonable, but still may not be optimal. Again, the optimizer is sensitive to the provided starting values and ocasionally returns NaNs due to pk values approaching 1 (which then results in a "divide by zero" problem when evaluating the Bernoulli likelihood). Despite this, the optimizer seems to converge to a reasonable solution:

$$\theta = \begin{bmatrix} .00358 & 1.5941 & 8.2981 \end{bmatrix}$$